

## REMARKS

Each issue the office action raises is addressed below after some general remarks

### **I. Objection to the Specification**

The embedded hyperlink has been deleted from the specification.

### **II. Claim 6**

Claim 6 has been amended to overcome the rejection.

### **III. 35 U.S.C. §112 ¶1 – Lack of enablement**

The invention identifies peptide motifs by comparing certain known peptide sequences. At a general level, identifying peptide sequences and comparing the same for similarity is well known. The instant invention is not simply directed to comparing sequences for similarity. Rather, in the instant invention, overlapping peptide sequences from several proteomes are compared. Claim 1 requires several steps. The following illustration is just one example of an embodiment of the invention and it is not provided to limit the claims.

The first step of the program is to generate a database of peptide sequences from different organisms. The database is generated from existing peptide databases, such as FASTA format protein files. The peptide sequences are broken down to overlapping peptides of specific length N, e.g., 8-mer (octapeptides). Redundant N-peptides are removed (See pages 13-14 of the specification) and the output is arranged to permit comparison, e.g., alphabetically according to single letter amino acid. This is carried out, e.g., using PEPLIB program. This can be done for different organisms and stored separately.

The second step is to use the output of PEPLIB from different organisms as input file and compare the overlapping redundant N-peptides using e.g., the program PEPLIMP.

Comparison of N-peptide sequences of different organisms will result in the identification of conserved N-peptides of the different organisms (See page 14 of the specification for details). As an illustration, comparison of Tb and Ec resulted in 5815 invariant N-peptides. Comparison of Tb, Ec and Bs peptide libraries resulted in 1767 invariant N-peptides.

Comparison of seven different organisms -- namely, Tb, Ec, Bs, Hp, Hi, Mg and Mp produced 164 common N-peptides (See page 14 for comparison tree).

Once the conserved N-peptide sequences are identified, the next step finds the location of these peptide sequences in the raw/original protein sequences and gives a protein identification number to the peptide sequences using the PEPXTRACT program (See page 15, Example 3).

The next step is to stitch together overlapping conserved peptide sequences, using the PEPSTITCH program. (See page 15).

Using the different programs mentioned above, 67 invariant peptides of different lengths were obtained on comparison of the seven different organisms (See page 8 of the Specification). Thus, this program can be used to identify conserved peptide sequences of any organisms available in the database. The program is not limited to the 7 organisms shown in the specification.

To cite an example, the peptide GIVGLPNVGKS (SEQ ID No. 22) is common to all the seven organisms compared (See page 15). More examples are listed on page 8 of the specification. Similarly, 6 invariant peptides were identified for the DNA gyrase protein which was possible using the computer programs of the instant invention. This protein is absent in human and is useful for drug targeting. Thus, the N-peptides identified within DNA gyrase protein are the structural determinants which can be used as a potential drug targets against bacterial infections. (See page 16 of the specification). The crystal structure of 3 of the conserved DNA gyrase peptides is shown in Figure 4. Thus, this program allows comparison of all microbial genomes for which protein data is available in the database.

Applicants note that this invention is correctly designated as bioinformatics. The steps of the invention are computations performed on databases. There is no need to "use" any biological organism and no compositions of matter or methods of treatment are claimed. The inventive program is not limited to any specific 7 organisms. As an example, we have compared only seven microbial proteomes with the human proteome, but the invention is not so limited.

Having reviewed the steps of the invention, we now study the information the Examiner concludes is lacking.

First (page 4, lines 1-4) the Examiner concludes that it is not clear whether the "common" peptides are indeed invariant or are merely conserved. The claims have now been amended to

Applicant : Samir Kumar Brahmachari et.al.  
Serial No. : 09/539,032  
Filed : March 30, 2000  
Page : 8

Attorney's Docket No.: 07064-010001 / 0317 - NF

overcome this rejection by consistently using the term "conserved". No distinction was intended by the use of other terms.

At page 4 lines 4-8, the examiner lists a number of steps that are taught in the specification. No response to those points is required.

At page 4, lines 8-page 5, line 9, the Examiner takes the position that the skill of the art is not adequate to compare genomic information to peptide sequences. Applicants do not concede this point, and they reserve their right to file additional claims in a continuation application raising the point again. The art is well aware how to use computational techniques to identify open reading frames and to infer peptide sequences from genomic information. Nevertheless, in order to expedite prosecution, the claims have been amended to overcome this rejection.

#### IV. 35 U.S.C. §112§2

Hyperlinks embedded in the claims have been deleted.

Amendments provide antecedent basis for the terms to which objection has been raised.

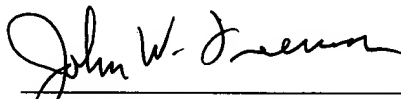
Other terms to which objection has been raised have been clarified.

Attached is a marked-up version of the changes being made by the current amendment.

Applicant asks that all claims be allowed. Enclosed is a \$930.00 check for the Petition for Extension of Time fee. Please apply any other charges or credits to Deposit Account No. 06-1050.

Respectfully submitted,

Date: 2/28/03

  
\_\_\_\_\_  
John W. Freeman, Esq.  
Reg. No. 29,066

Fish & Richardson P.C.  
225 Franklin Street  
Boston, Massachusetts 02110-2804  
Telephone: (617) 542-5070  
Facsimile: (617) 542-8906

**Version with markings to show changes made**

Claims 1-9 have been amended as follows:

1. A computer-based method for identifying [invariant] conserved peptide motifs useful as drug targets wherein the said method comprises the steps of:
  - i) providing electronic data representing [generating computationally overlapping] peptide libraries from [all] the protein sequences of [the] selected organisms [available at <http://www.ncbi.nlm.nih.gov>],
  - ii) from the data of step (i), generating computationally overlapping peptide sequences of length 'N', and sorting computationally the peptide[s] sequences of length 'N' [obtained as above, alphabetically,] according to [single letter] amino acid [code] sequence,
  - iii) matching computationally the sorted peptide sequences of length 'N' [common peptide sequences] of the selected [bacteria] organisms to produce matched common peptide sequences,
  - iv) locating computationally [these] the matched common peptide[s] sequences in the [original] protein[s] sequences of step i) and subsequently labeling [them] the matched common peptide sequences with their origin and location,
  - v) joining computationally [the] overlapping common peptide[s] sequences to obtain [a long chain of invariant] extended conserved peptide sequences,
  - vi) annotating secondary structure of [these] extended conserved peptide[s] sequences [from the] based on a crystal structure database,
  - vii) comparing pathogenic strain [genomes] proteomes against [genomes] proteomes of non-pathogenic strains and selecting [the] at least one conserved peptide sequence[s] not commonly conserved in these two groups,
  - viii) validating computationally [the] at least one conserved peptide [invariant] sequence [motifs] as a potential drug target sequence by searching for [the] a given conserved sequences in the host [genome] proteome and rejecting [the ones] sequences present in the host [genome] proteome.
2. The method of claim 1 wherein [the length of the sliding window of length 'N' ranges from 4 to any length of amino acid residues] 'N' is at least 4.

3. The method of claim 1 wherein the [protein sequence data is taken from any organism but not specifically limited to microbes such as] selected organisms include at least one of: Mycoplasma pneumoniae, Helicobacter pylori, Hemophilus influenzae, Mycobacterium tuberculosis, Mycoplasma genitalium, Bacillus subtilis, Escherichia coli.

4. A method as claimed in claim 1 where conserved peptide sequences include one or more of [motifs as modified comprising]:

1.	AAQSIGEPGTQLT	35.	KMSKSKGN
2.	AGDGTTTAT	36.	KMSKSLGN
3.	AGRHGKNG	37.	KNMITGAAQMDGAILVV
4.	AHIDAGKTTT	38.	KPNSALRK
5.	CPIETPEG	39.	LFGGAGVGKTV
6.	DEPSIGLH	40.	LGPSGCGK
7.	DEPTSALD	41.	LHAGGKFD
8.	DEPTTALDVT	42.	LIDEARTPLIISG
9.	DHAGIATQ	43.	LLNRAPTLH
10.	DHPHGGGEG	44.	LPDKAIDLIDE
11.	DLGGGTFD	45.	LPGKLADC
12.	DVLDTWFSS	46.	LSGGQQQR
13.	ERERGITI	47.	MGHVDHGKT
14.	ERGITITSAAT	48.	NADFDGDQMAVH
15.	ESRRIDNQLRGR	49.	NGAGKSTL
16.	FSGGQRQR	50.	NLLGKRVD
17.	GEPGVGKTA	51.	NTDAEGRL
18.	GFDYL RDN	52.	PSAVGYQPTLA
19.	GHNLQEHS	53.	QRVALARA
20.	GIDL GTTNS	54.	QRYKGLGEM
21.	GINLLREGLD	55.	RDGLKPVHRR
22.	GIVGLPNVGKS	56.	SALDVSIQA
23.	GKSSLLNA	57.	SGGLHGVG
24.	GLTGRKIIVD TYG	58.	SGSGKSSL
25.	GPPGTGKTLLA	59.	SGSGKSTL
26.	GPPGVGKT	60.	SVFAGVGERTREGND
27.	GSGKTLL	61.	TGRTHQIRVH
28.	GTRIFGPV	62.	TGVSGSGKS
29.	IDTPGHVDFT	63.	TLSGGEAQRI

30.	ILAHIDHGKSTL	64.	TNKYAEGYP
31.	INGFGRIGR	65.	TPRSNPATY
32.	IREGGRTVG	66.	VEGDSAGG
33.	IVGESGSGKS	67.	VRKRPGMYIG
34.	KFSTYATWWI		

5. A method as claimed in claim 1 [wherein] comprising increasing the number of [invariant] conserved peptide[s] sequences [varies according to the relatedness] by increasing the relatedness among the organisms [and the number of organisms] being compared.

6. A method as claimed in any one of claims 1-4 wherein the invariant sequences belong to at least one of the following proteins [as available in the database <http://www.ncbi.nlm.nih.gov> wherein the said list of proteins comprise]:

I DNA DIRECTED RNA POLYMERASE BETA CHAIN

II EXCINUCLEASE ABC SUBUNIT A

III EXCINUCLEASE ABC SUBUNIT B

IV DNA GYRASE SUBUNIT B

V ATP SYNTHASE BETA CHAIN

VI S-ADENOSYLMETHIONINE SYNTHETASE

VII GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE

VIII ELONGATION FACTOR G (EF-G)

IX ELONGATION FACTOR TU (EF-TU)

X 30S RIBOSOMAL PROTEIN S12

XI 50S RIBOSOMAL PROTEIN L12

XII 50S RIBOSOMAL PROTEIN L14

XIII VALYL tRNA SYNTHETASE (VALRS)

XIV CELL DIVISION PROTEIN FtsH HOMOLOG

XV DnaK PROTEIN (HSP70)

XVI GTP BINDING PROTEIN LepA

XVII TRANSPORTER

XVIII OLIGOPEPTIDE TRANSPORT ATP BINDING PROTEIN OPPF

7. A method as claimed in claim 1 wherein the said method of comparing the peptide libraries as given in step (iii) of claim 1 is carried out by following the steps [given in figure 1]:

- selecting organism names from a menu;
- iteratively comparing peptide sequences of a first organism to peptide sequences of a second organism and for matching sequences, writing sequences to a file for the first organism and to a file for the second organism.

8. A method as claimed in claim 1 wherein the said method of locating the common peptides in the original protein sequences as given in step (iv) of claim 1 is carried out by following the steps [given in figure 2]:

- selecting protein sequences;
- iteratively comparing matched peptide sequences to protein sequences;
- where the peptide exists in a protein sequence writing the peptide PID, location and organism in a file associated with that peptide.

9. A method as claimed in claim 1 wherein the said method of creating a common peptide of variable length after removing the overlapping as given in step (v) of claim 1 is carried out by following the steps [given in figure 3]:

- iteratively comparing data on matched peptide locations;
- determining overlapping matched peptides; and
- determining extended peptide sequences based on overlapping matched peptide sequences.